Mathematical Expectation: Simple Random Variables*

Paul E Pfeiffer

Abstract

For simple, real valued random variables, the expectation is the probability weighted average of the values taken on. It may be viewed as the center of mass for the probability mass distribution on the line.

1 Introduction

The probability that real random variable X takes a value in a set M of real numbers is interpreted as the likelihood that the observed value $X(\omega)$ on any trial will lie in M. Historically, this idea of likelihood is rooted in the intuitive notion that if the experiment is repeated enough times the probability is approximately the fraction of times the value of X will fall in M. Associated with this interpretation is the notion of the *average* of the values taken on. We incorporate the concept of *mathematical expectation* into the mathematical model as an appropriate form of such averages. We begin by studying the mathematical expectation of simple random variables, then extend the definition and properties to the general case. In the process, we note the relationship of mathematical expectation to the Lebesque integral, which is developed in abstract measure theory. Although we do not develop this theory, which lies beyond the scope of this study, identification of this relationship provides access to a rich and powerful set of properties which have far reaching consequences in both application and theory.

2 Expectation for simple random variables

The notion of mathematical expectation is closely related to the idea of a weighted mean, used extensively in the handling of numerical data. Consider the arithmetic average $\overline{x}$ of the following ten numbers: 1, 2, 2, 2, 4, 5, 5, 8, 8, 8, which is given by

$$\overline{x} = \frac{1}{10}(1+2+2+2+4+5+5+8+8+8) \tag{1}$$

Examination of the ten numbers to be added shows that five distinct values are included. One of the ten, or the fraction 1/10 of them, has the value 1, three of the ten, or the fraction 3/10 of them, have the value 2, 1/10 has the value 4, 2/10 have the value 5, and 3/10 have the value 8. Thus, we could write

$$\overline{x} = (0.1 \cdot 1 + 0.3 \cdot 2 + 0.1 \cdot 4 + 0.2 \cdot 5 + 0.3 \cdot 8) \tag{2}$$

*Version 1.6: Sep 18, 2009 1:17 pm -0500

The pattern in this last expression can be stated in words: Multiply each possible value by the fraction of the numbers having that value and then sum these products. The fractions are often referred to as the *relative frequencies*. A sum of this sort is known as a *weighted average*.

In general, suppose there are n numbers $\{x_1,\ x_2,\ \cdots\ x_n\}$ to be averaged, with $m \le n$ distinct values $\{t_1,\ t_2,\ \cdots,\ t_m\}$. Suppose f_1 have value t_1, f_2 have value t_2, $\cdots$, f_m have value t_m. The f_i must add to n. If we set $p_i = f_i/n$, then the fraction p_i is called the relative frequency of those numbers in the set which have the value $t_i, 1 \le i \le m$. The average $\overline{x}$ of the n numbers may be written

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \sum_{j=1}^{m} t_j p_j \qquad (3)$$

In probability theory, we have a similar averaging process in which the relative frequencies of the various possible values of are replaced by the probabilities that those values are observed on any trial.

Definition. For a simple random variable X with values $\{t_1,\ t_2,\ \cdots,\ t_n\}$ and corresponding probabilities $p_i = P(X = t_i)$, the *mathematical expectation*, designated $E[X]$, is the probability weighted average of the values taken on by X. In symbols

$$E[X] = \sum_{i=1}^{n} t_i P(X = t_i) = \sum_{i=1}^{n} t_i p_i \qquad (4)$$

Note that the expectation is determined by the distribution. Two quite different random variables may have the same distribution, hence the same expectation. Traditionally, this average has been called the *mean*, or the *mean value*, of the random variable X.

Example 1: Some special cases

1. Since $X = aI_E = 0I_{E^c} + aI_E$, we have $E[aI_E] = aP(E)$.
2. For X a constant c, $X = cI_\Omega$, so that $E[c] = cP(\Omega) = c$.
3. If $X = \sum_{i=1}^{n} t_i I_{A_i}$ then $aX = \sum_{i=1}^{n} at_i I_{A_i}$, so that

$$E[aX] = \sum_{i=1}^{n} at_i P(A_i) = a\sum_{i=1}^{n} t_i P(A_i) = aE[X] \qquad (5)$$

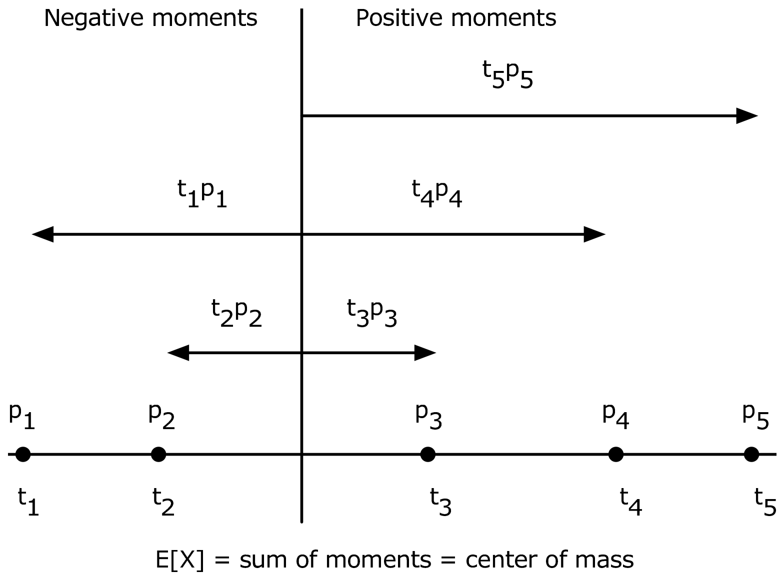

Figure 1: Moment of a probability distribution about the origin.

Mechanical interpretation

In order to aid in visualizing an essentially abstract system, we have employed the notion of probability as mass. The distribution induced by a real random variable on the line is visualized as a unit of probability mass actually distributed along the line. We utilize the mass distribution to give an important and helpful mechanical interpretation of the expectation or mean value. In Example 6[1] in "Mathematical Expectation: General Random Variables", we give an alternate interpretation in terms of mean-square estimation.

Suppose the random variable X has values $\{t_i : 1 \leq i \leq n\}$, with $P(X = t_i) = p_i$. This produces a probability mass distribution, as shown in Figure 1, with point mass concentration in the amount of p_i at the point t_i. The expectation is

$$\sum_i t_i p_i \tag{6}$$

Now $|t_i|$ is the distance of point mass p_i from the origin, with p_i to the left of the origin iff t_i is negative. Mechanically, the sum of the products $t_i p_i$ is the *moment* of the probability mass distribution about the origin on the real line. From physical theory, this moment is known to be the same as the product of the total mass times the number which locates the center of mass. Since the total mass is one, the *mean value* is the *location of the center of mass*. If the real line is viewed as a stiff, weightless rod with point mass p_i attached at each value t_i of X, then the mean value μ_X is the point of balance. Often there are symmetries in the distribution which make it possible to determine the expectation without detailed calculation.

Example 2: The number of spots on a die

Let X be the number of spots which turn up on a throw of a simple six-sided die. We suppose each number is equally likely. Thus the values are the integers one through six, and each probability is

[1] "Mathematical Expectation; General Random Variables", Example 6: Alternate interpretation of the mean value <http://cnx.org/content/m23412/latest/#fs-id7202349>

1/6. By definition

$$E[X] = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} \tag{7}$$

Although the calculation is very simple in this case, it is really not necessary. The probability distribution places equal mass at each of the integer values one through six. The center of mass is at the midpoint.

Example 3: A simple choice

A child is told she may have one of four toys. The prices are \$2.50. \$3.00, \$2.00, and \$3.50, respectively. She choses one, with respective probabilities 0.2, 0.3, 0.2, and 0.3 of choosing the first, second, third or fourth. What is the expected cost of her selection?

$$E[X] = 2.00 \cdot 0.2 + 2.50 \cdot 0.2 + 3.00 \cdot 0.3 + 3.50 \cdot 0.3 = 2.85 \tag{8}$$

For a simple random variable, the mathematical expectation is determined as the dot product of the value matrix with the probability matrix. This is easily calculated using MATLAB.

Example 4: MATLAB calculation for Example 3

```
    X = [2 2.5 3 3.5];  % Matrix of values (ordered)
PX = 0.1*[2 2 3 3]; % Matrix of probabilities
EX = dot(X,PX)      % The usual MATLAB operation
EX =  2.8500
Ex = sum(X.*PX)     % An alternate calculation
Ex =  2.8500
ex = X*PX'          % Another alternate
ex =  2.8500
```

Expectation and primitive form

The definition and treatment above assumes X is in *canonical form*, in which case

$$X = \sum_{i=1}^{n} t_i I_{A_i}, \quad \text{where } A_i = \{X = t_i\}, \text{ implies } E[X] = \sum_{i=1}^{n} t_i P(A_i) \tag{9}$$

We wish to ease this restriction to canonical form.

Suppose simple random variable X is in a *primitive form*

$$X = \sum_{j=1}^{m} c_j I_{C_j}, \text{ where } \{C_j : 1 \le j \le m\} \text{ is a partition} \tag{10}$$

We show that

$$E[X] = \sum_{j=1}^{m} c_j P(C_j) \tag{11}$$

Before a formal verification, we begin with an example which exhibits the essential pattern. Establishing the general case is simply a matter of appropriate use of notation.

Example 5: Simple random variable X in primitive form

$$X = I_{C_1} + 2I_{C_2} + I_{C_3} + 3I_{C_4} + 2I_{C_5} + 2I_{C_6}, \quad \text{with } \{C_1, C_2, C_3, C_4, C_5. C_6\} \text{ a partition} \tag{12}$$

Inspection shows the distinct possible values of X to be 1, 2, or 3. Also,

$$A_1 = \{X = 1\} = C_1 \bigvee C_3, \quad A_2 = \{X = 2\} = C_2 \bigvee C_5 \bigvee C_6 \text{ and } A_3 = \{X = 3\} = C_4 \tag{13}$$

so that

$$P(A_1) = P(C_1) + P(C_3), \quad P(A_2) = P(C_2) + P(C_5) + P(C_6), \text{ and } P(A_3) = P(C_4) \tag{14}$$

Now

$$E[X] = P(A_1) + 2P(A_2) + 3P(A_3) = P(C_1) + P(C_3) + 2[P(C_2) + P(C_5) + P(C_6)] + 3P(C_4) \tag{15}$$

$$= P(C_1) + 2P(C_2) + P(C_3) + 3P(C_4) + 2P(C_5) + 2P(C_6) \tag{16}$$

To establish the general pattern, consider $X = \sum_{j=1}^{m} c_j I_{C_j}$. We identify the distinct set of values contained in the set $\{c_j : 1 \leq j \leq m\}$. Suppose these are $t_1 < t_2 < \cdots < t_n$. For any value t_i in the range, identify the index set J_i of those j such that $c_j = t_i$. Then the terms

$$\sum_{J_i} c_j I_{C_j} = t_i \sum_{J_i} I_{C_j} = t_i I_{A_i}, \quad \text{where } A_i = \bigvee_{j \in J_i} C_j \tag{17}$$

By the additivity of probability

$$P(A_i) = P(X = t_i) = \sum_{j \in J_i} P(C_j) \tag{18}$$

Since for each $j \in J_i$ we have $c_j = t_i$, we have

$$E[X] = \sum_{i=1}^{n} t_i P(A_i) = \sum_{i=1}^{n} t_i \sum_{j \in J_i} P(C_j) = \sum_{i=1}^{n} \sum_{j \in J_i} c_j P(C_j) = \sum_{j=1}^{m} c_j P(C_j) \tag{19}$$

— □

Thus, *the defining expression for expectation thus holds for X in a primitive form.*

An alternate approach to obtaining the expectation from a primitive form is to use the csort operation to determine the distribution of X from the coefficients and probabilities of the primitive form.

Example 6: Alternate determinations of $E[X]$

Suppose X in a primitive form is

$$X = I_{C_1} + 2I_{C_2} + I_{C_3} + 3I_{C_4} + 2I_{C_5} + 2I_{C_6} + I_{C_7} + 3I_{C_8} + 2I_{C_9} + I_{C_{10}} \tag{20}$$

with respective probabilities

$$P(C_i) = 0.08,\ 0.11,\ 0.06,\ 0.13,\ 0.05,\ 0.08,\ 0.12,\ 0.07,\ 0.14,\ 0.16 \tag{21}$$

```
    c = [1 2 1 3 2 2 1 3 2 1];            % Matrix of coefficients
pc = 0.01*[8 11 6 13 5 8 12 7 14 16]; % Matrix of probabilities
EX = c*pc'
EX =   1.7800                         % Direct solution
[X,PX] = csort(c,pc);                 % Determination of dbn for X
disp([X;PX]')
    1.0000    0.4200
```

```
    2.0000    0.3800
    3.0000    0.2000
Ex = X*PX'                              %  E[X] from distribution
Ex =   1.7800
```

Linearity

The result on primitive forms may be used to establish the *linearity* of mathematical expectation for simple random variables. Because of its fundamental importance, we work through the verification in some detail.

Suppose $X = \sum_{i=1}^{n} t_i I_{A_i}$ and $Y = \sum_{j=1}^{m} u_j I_{B_j}$ (both in canonical form). Since

$$\sum_{i=1}^{n} I_{A_i} = \sum_{j=1}^{m} I_{B_j} = 1 \quad (22)$$

we have

$$X + Y = \sum_{i=1}^{n} t_i I_{A_i} \left(\sum_{j=1}^{m} I_{B_j} \right) + \sum_{j=1}^{m} u_j I_{B_j} \left(\sum_{i=1}^{n} I_{A_i} \right) = \sum_{i=1}^{n} \sum_{j=1}^{m} (t_i + u_j) I_{A_i} I_{B_j} \quad (23)$$

Note that $I_{A_i} I_{B_j} = I_{A_i B_j}$ and $A_i B_j = \{X = t_i, Y = u_j\}$. The class of these sets for all possible pairs (i, j) forms a partition. Thus, the last summation expresses $Z = X + Y$ in a primitive form. Because of the result on primitive forms, above, we have

$$E[X + Y] = \sum_{i=1}^{n} \sum_{j=1}^{m} (t_i + u_j) P(A_i B_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} t_i P(A_i B_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} u_j P(A_i B_j) \quad (24)$$

$$= \sum_{i=1}^{n} t_i \sum_{j=1}^{m} P(A_i B_j) + \sum_{j=1}^{m} u_j \sum_{i=1}^{n} P(A_i B_j) \quad (25)$$

We note that for each i and for each j

$$P(A_i) = \sum_{j=1}^{m} P(A_i B_j) \quad \text{and} \quad P(B_j) = \sum_{i=1}^{n} P(A_i B_j) \quad (26)$$

Hence, we may write

$$E[X + Y] = \sum_{i=1}^{n} t_i P(A_i) + \sum_{j=1}^{m} u_j P(B_j) = E[X] + E[Y] \quad (27)$$

Now aX and bY are simple if X and Y are, so that with the aid of Example 1 (Some special cases) we have

$$E[aX + bY] = E[aX] + E[bY] = aE[X] + bE[Y] \quad (28)$$

If X, Y, Z are simple, then so are $aX + bY$, and cZ. It follows that

$$E[aX + bY + cZ] = E[aX + bY] + cE[Z] = aE[X] + bE[Y] + cE[Z] \quad (29)$$

By an inductive argument, this pattern may be extended to a linear combination of any finite number of simple random variables. Thus we may assert

Linearity. The expectation of a linear combination of a finite number of simple random variables is that linear combination of the expectations of the individual random variables.

— □

Expectation of a simple random variable in affine form

As a direct consequence of linearity, whenever simple random variable X is in affine form, then

$$E[X] = E\left[c_0 + \sum_{i=1}^{n} c_i I_{E_i}\right] = c_0 + \sum_{i=1}^{n} c_i P(E_i) \tag{30}$$

Thus, the defining expression holds for any affine combination of indicatorfunctions, whether in canonical form or not.

Example 7: Binomial distribution (n, p)

This random variable appears as the number of successes in n Bernoulli trials with probability p of success on each component trial. It is naturally expressed in affine form

$$X = \sum_{i=1}^{n} I_{E_i} \quad \text{so that} \quad E[X] = \sum_{i=1}^{n} p = np \tag{31}$$

Alternately, in canonical form

$$X = \sum_{k=0}^{n} k I_{A_{kn}}, \quad \text{with} \quad p_k = P(A_{kn}) = P(X = k) = C(n, k) p^k q^{n-k}, \quad q = 1 - p \tag{32}$$

so that

$$E[X] = \sum_{k=0}^{n} k C(n, k) p^k q^{n-k}, \quad q = 1 - p \tag{33}$$

Some algebraic tricks may be used to show that the second form sums to np, but there is no need of that. The computation for the affine form is much simpler.

Example 8: Expected winnings

A bettor places three bets at \$2.00 each. The first bet pays \$10.00 with probability 0.15, the second pays \$8.00 with probability 0.20, and the third pays \$20.00 with probability 0.10. What is the expected gain?

SOLUTION

The net gain may be expressed

$$X = 10 I_A + 8 I_B + 20 I_C - 6, \quad \text{with} \quad P(A) = 0.15, \; P(B) = 0.20, \; P(C) = 0.10 \tag{34}$$

Then

$$E[X] = 10 \cdot 0.15 + 8 \cdot 0.20 + 20 \cdot 0.10 - 6 = -0.90 \tag{35}$$

These calculations may be done in MATLAB as follows:

```
    c = [10 8 20 -6];
p = [0.15 0.20 0.10 1.00]; % Constant a = aI_(Omega), with P(Omega) = 1
E = c*p'
E =  -0.9000
```

Functions of simple random variables

If X is in a primitive form (including canonical form) and g is a real function defined on the range of X, then

$$Z = g(X) = \sum_{j=1}^{m} g(c_j) I_{C_j} \quad \text{a primitive form} \tag{36}$$

so that

$$E[Z] = E[g(X)] = \sum_{j=1}^{m} g(c_j) P(C_j) \tag{37}$$

Alternately, we may use csort to determine the distribution for Z and work with that distribution.

Caution. If X is in affine form (but not a primitive form)

$$X = c_0 + \sum_{j=1}^{m} c_j I_{E_j} \quad \text{then} \quad g(X) \neq g(c_0) + \sum_{j=1}^{m} g(c_j) I_{E_j} \tag{38}$$

so that

$$E[g(X)] \neq g(c_0) + \sum_{j=1}^{m} g(c_j) P(E_j) \tag{39}$$

Example 9: Expectation of a function of X

Suppose X in a primitive form is

$$X = -3I_{C_1} - I_{C_2} + 2I_{C_3} - 3I_{C_4} + 4I_{C_5} - I_{C_6} + I_{C_7} + 2I_{C_8} + 3I_{C_9} + 2I_{C_{10}} \tag{40}$$

with probabilities $P(C_i) = 0.08,\ 0.11,\ 0.06,\ 0.13,\ 0.05,\ 0.08,\ 0.12,\ 0.07,\ 0.14,\ 0.16$.

Let $g(t) = t^2 + 2t$. Determine $E[g(X)]$.

```
   c = [-3 -1 2 -3 4 -1 1 2 3 2];            % Original coefficients
pc = 0.01*[8 11 6 13 5 8 12 7 14 16];     % Probabilities for C_j
G = c.^2 + 2*c                            % g(c_j)
G =  3  -1   8   3  24  -1   3   8  15   8
EG = G*pc'                                % Direct computation
EG =  6.4200
[Z,PZ] = csort(G,pc);                     % Distribution for Z = g(X)
disp([Z;PZ]')                             % Optional display
   -1.0000    0.1900
    3.0000    0.3300
    8.0000    0.2900
   15.0000    0.1400
   24.0000    0.0500
EZ = Z*PZ'                                % E[Z] from distribution for Z
EZ =  6.4200
```

A similar approach can be made to a function of a pair of simple random variables, provided the joint distribution is available. Suppose $X = \sum_{i=1}^{n} t_i I_{A_i}$ and $Y = \sum_{j=1}^{m} u_j I_{B_j}$ (both in canonical form). Then

$$Z = g(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} g(t_i, u_j) I_{A_i B_j} \tag{41}$$

The $A_i B_j$ form a partition, so Z is in a primitive form. We have the same two alternative possibilities: (1) direct calculation from values of $g(t_i, u_j)$ and corresponding probabilities $P(A_i B_j) = P(X = t_i, Y = u_j)$, or (2) use of csort to obtain the distribution for Z.

Example 10: Expectation for $Z = g(X, Y)$

We use the joint distribution in file jdemo1.m and let $g(t, u) = t^2 + 2tu - 3u$. To set up for calculations, we use jcalc.

```
    % file jdemo1.m
X = [-2.37 -1.93 -0.47 -0.11  0  0.57 1.22 2.15 2.97 3.74];
Y = [-3.06 -1.44 -1.21  0.07 0.88 1.77 2.01 2.84];
P = 0.0001*[ 53   8 167 170 184  18  67 122  18  12;
             11  13 143 221 241 153  87 125 122 185;
            165 129 226 185  89 215  40  77  93 187;
            165 163 205  64  60  66 118 239  67 201;
            227   2 128  12 238 106 218 120 222  30;
             93  93  22 179 175 186 221  65 129   4;
            126  16 159  80 183 116  15  22 113 167;
            198 101 101 154 158  58 220 230 228 211];
```

```
    jdemo1                   % Call for data
jcalc                     % Set up
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
G = t.^2 + 2*t.*u - 3*u; % Calculation of matrix of [g(t_i, u_j)]
EG = total(G.*P)         % Direct calculation of expectation
EG =  3.2529
[Z,PZ] = csort(G,P);     % Determination of distribution for Z
EZ = Z*PZ'               % E[Z] from distribution
EZ =  3.2529
```